PUBLICATION

# EXAM-TM50H-POST


## POST-TEST — SL 5H: ADVANCED AI ENGINEER

*Maven Smart System (MSS) — USAREUR-AF*

HEADQUARTERS
UNITED STATES ARMY EUROPE AND AFRICA
(USAREUR-AF)
Wiesbaden, Germany



**26 MARCH 2026**

DRAFT — UNOFFICIAL — NOT FOR OPERATIONAL USE

# POST-TEST — SL 5H: ADVANCED AI ENGINEER

## MAVEN SMART SYSTEM (MSS) — USAREUR-AF

| Field | Detail |
|---|---|
| **Course** | SL 5H: Advanced AI Engineer |
| **Form** | Post-Test |
| **Level** | SL 5H (Advanced Specialist) |
| **Audience** | Senior AI engineers; prerequisite: SL 4H + production AI experience |
| **Time Allowed** | 45 minutes |
| **Passing Score** | 70% (42/60) |

## INSTRUCTIONS

This assessment evaluates mastery of course learning objectives. A passing score of 70% is required to receive credit. Complete independently without reference to training materials.

## SECTION 1 — MULTIPLE CHOICE

*Circle the letter of the best answer. (2 points each)*

**1. A multi-agent system has three agents: a retrieval agent, a generation agent, and a validation agent. The retrieval agent begins timing out intermittently. Without a circuit breaker, the most likely consequence is:**

A. Generation agent receives empty or partial context, produces low-quality or hallucinated output, and the bad output propagates through the validation step — corrupting the Ontology with incorrect data B. The validation agent will catch the errors and produce correct output despite the retrieval failures C. The

system automatically retries the retrieval agent until success D. The generation agent skips retrieval context and falls back to its base model knowledge

**2. The rate-limiting requirement for an agent chain that can write to the Foundry Ontology is specifically designed to:**

A. Prevent an uncontrolled agent chain from executing a large number of irreversible Ontology writes before a human can detect and intervene in a failure scenario B. Comply with Foundry API rate limits to prevent HTTP 429 errors C. Ensure the agent does not exceed compute budget allocated for the project D. Limit agent outputs to the number of records a human reviewer can process per shift

**3. Fine-tuning an LLM on Army operational corpora (lessons learned, SITREPs, intelligence summaries) requires SJA review because:**

A. All AI model training on Army networks requires SJA review under Army Regulation B. The corpus may contain operationally sensitive information that would be encoded into model weights and potentially accessible via adversarial prompting of the fine-tuned model in other contexts C. SJA must confirm the model does not constitute a weapon system under DoD 5000.01 D. The review determines the classification level of the fine-tuned model artifact

**4. In a hybrid retrieval RAG architecture, combining dense and sparse retrieval improves over dense-only retrieval because:**

A. Dense retrieval alone is too computationally expensive for production deployments B. Dense retrieval does not support chunked documents C. Sparse retrieval enforces classification constraints that dense retrieval cannot D. Sparse retrieval captures exact-match keywords (military unit designators, equipment NSNs, location names) that semantic embeddings may fail to retrieve accurately

**5. A corpus quality evaluation for a RAG system used for operational planning should assess:**

A. Document count only — larger corpora always produce better retrieval quality B. Whether documents are in PDF or plain-text format C. The embedding model's performance on a generic benchmark dataset D. Document freshness, deduplication, classification handling, coverage of relevant topics, and chunk coherence — poor corpus quality directly degrades generation quality regardless of LLM capability

**6. Adversarial prompt injection testing MUST occur in an isolated environment because:**

A. Army policy prohibits testing adversarial techniques on production systems B. Production inference endpoints have rate limits that would prevent effective testing C. Injected adversarial prompts could cause the system to write incorrect data to the production Ontology, send unauthorized communications, or expose sensitive information to the attacker D. Isolated environments provide a ground-truth dataset to evaluate injection success rates

**7. A SL 5H AI observability pipeline detects that the acceptance rate of AI-generated outputs by human reviewers has dropped from 87% to 61% over four weeks. The correct response is:**

A. Treat this as a quality degradation alert — investigate root cause (input distribution shift, prompt drift, model issue), remediate before the rate drops further, and consider rollback if the rate approaches the minimum acceptable threshold B. Increase the frequency of human review to catch more errors C. The drop is within normal operational variance — continue monitoring D. Retrain the model on the last four weeks of rejected outputs to improve performance

**8. Per DoD RAIMTF guidelines (2024), "governability" in an AI system means:**

A. The system must be operated under a signed MOU between the operating command and JAIC B. The AI system must be able to be modified, retrained, shut down, or have its decisions overridden by authorized humans C. A governance board must approve all AI outputs before they are used operationally D. The system's decision logic must be approved by the chain of command annually

**9. The Army CIO Memorandum (April 2024) assigns responsibility for AI-generated content to:**

A. The AI system developer, who must certify accuracy before the system is deployed B. The commanding officer of the unit that deployed the AI system C. The human reviewer who reviewed and approved the content for official use — AI does not bear legal responsibility D. The data steward who manages the AI system's input data

**10. An enterprise AI architecture review for a new multi-modal AI system (text + imagery) must verify which classification constraint?**

A. The system must be deployed on SIPR regardless of the data it processes B. The inference endpoint's classification level must be appropriate for the highest classification of any input modality — if imagery is handled at a higher level than text, the endpoint must operate at the imagery classification level C. Multi-modal systems require separate endpoints for each input modality D. Imagery inputs must be downsampled to reduce classification risk before processing

**11. A production AI system's monitoring pipeline should alert when which of the following is detected?**

A. The model produces an output containing a word not in its training vocabulary B. The model produces outputs longer than the average output length from the previous week C. A single human reviewer disagrees with two model outputs in one shift D. Consecutive human reviewer rejections exceed a defined threshold, the input PSI crosses 0.20, or latency p95 exceeds the SLA

**12. An agent configured with both a document retrieval tool and a Foundry Ontology write tool should require human approval before executing the write tool because:**

A. Ontology writes require Editor role which agents do not normally hold B. The Foundry API does not support agent-initiated write transactions C. Ontology writes are irreversible — incorrect writes can corrupt operational data, and human approval ensures accountability and prevents runaway agent writes D. Write operations exceed the inference endpoint's token limit

**13. "Corpus quality drift" in a RAG system refers to:**

A. The embedding model's performance degrading over time due to model aging B. The corpus of retrievable documents becoming outdated, duplicated, or misaligned with current operational context, causing retrieval to return stale or irrelevant content C. The chunking strategy becoming misaligned with document lengths as new documents are added D. Classification labels on corpus documents becoming incorrect over time

**14. A SL 5H AI governance checklist requires that all enterprise AI systems have:**

A. A separate MSS project for each AI model in production B. Quarterly GO-level review of all AI outputs C. Documented ownership, a defined review cycle, a decommission plan, and compliance mapping to Army CIO Memo and DoD RAIMTF requirements D. A dedicated AI engineer with SL 5H qualification assigned to each system

**15. The most important security control for an enterprise AI system that processes operational data is:**

A. Defense-in-depth: classification-level-appropriate inference endpoint, access controls on input data, sanitized prompts, output review, and monitoring — no single control is sufficient alone B. Using a private LLM that does not send data outside the government network C. Encrypting all prompts before sending to the inference endpoint D. Restricting system access to users with a minimum TS/SCI clearance

---

## SECTION 2 — SHORT ANSWER

*Answer in 2–5 sentences. (6 points each)*

**SA-1. Design the multi-agent architecture for a system that: (1) monitors incoming field reports, (2) extracts key entities, (3) generates an operational summary, and (4) routes for human review. For each agent, describe its function, input/output, and the circuit breaker and rate-limiting controls you would implement.**

**SA-2. The USAREUR-AF AI engineering team is evaluating whether to use RAG or fine-tuning for a new system that will answer questions about current Theater operations. Make the architectural decision and defend it, citing the classification constraints and operational tempo factors that drive your recommendation.**

**SA-3. You are leading a red-team assessment of a production AIP Logic workflow that processes maintenance SITREPs. Describe your complete red-team methodology: what adversarial scenarios you test, the environment where testing occurs, how you document results, and the threshold for release versus remediation.**

**SA-4. Describe the enterprise AI governance framework you would design for a theater-level AI deployment. Include: ownership structure, review cycle, compliance mapping, and what happens when the system fails its governance review.**

**SA-5. Six months after deploying a multi-agent system in production, you receive a report that the system wrote 47 incorrect maintenance records to the Ontology over a three-day period before anyone noticed. Conduct a root-cause analysis: what failure controls should have caught this, why they failed, and what architectural changes would prevent recurrence.**



## SCORING SUMMARY

| Section | Questions | Points Each | Total Points |
|---|---|---|---|
| Multiple Choice | 15 | 2 | 30 |
| Short Answer | 5 | 6 | 30 |

| Section | Questions | Points Each | Total Points |
|---|---|---|---|
| **Total** | — | — | **60** |

Passing: 42/60 (70%) — Post-test only. Pre-test is diagnostic.

## ANSWER KEY — INSTRUCTOR USE ONLY

*Do not distribute to students.*

**Multiple Choice:** 1. A — Without circuit breaker, partial/empty context propagates through the chain producing bad outputs that corrupt the Ontology. 2. A — Rate limiting bounds irreversible Ontology writes before human intervention is possible. 3. B — Operational corpus encodes sensitive information into weights accessible via adversarial prompting. 4. D — Sparse retrieval captures exact-match military identifiers (NSNs, unit designators) that semantic embedding misses. 5. D — Corpus quality must cover freshness, deduplication, classification, topic coverage, and chunk coherence. 6. C — Injected prompts in production could write incorrect data, send unauthorized comms, or expose sensitive information. 7. A — Quality degradation alert requires root cause investigation and remediation; consider rollback. 8. B — Governability = ability to modify, retrain, shut down, or override by authorized humans. 9. C — Human reviewer who approved AI content bears responsibility per Army CIO Memo. 10. B — Endpoint classification level must match the highest classification of any input modality. 11. D — Alert on consecutive reviewer rejections exceeding threshold, PSI > 0.20, or latency p95 SLA breach. 12. C — Ontology writes are irreversible; human approval prevents runaway writes and ensures accountability. 13. B — Corpus quality drift = documents becoming outdated, duplicated, or misaligned with operational context. 14. C — Enterprise AI governance requires ownership, review cycle, decommission plan, and compliance mapping. 15. A — Defense-in-depth: no single control is sufficient; classification endpoint + access controls + output review + monitoring.

**Short Answer Guidance:**

SA-1. Full credit: Agent 1 (Monitor): reads new field reports from Ontology; circuit breaker: if no new reports in expected window, alert rather than fail; Agent 2 (Entity Extractor): extracts entities in structured JSON; circuit breaker: if JSON output malformed or confidence below threshold, stop chain, log error; Agent 3 (Summary Generator): generates operational summary from entities; rate limit: max N summaries per hour to match human review capacity; circuit breaker: if quality check fails, route to error queue not production; Agent 4 (Router): places summary in Draft status in human review queue; never auto-publishes. All four agents, their I/O, and circuit breaker/rate control for each required.

SA-2. Full credit: recommendation — RAG, not fine-tuning; reasons: (1) Theater operations data changes daily — fine-tuning cannot keep pace without continuous expensive retraining; (2) operational SITREPs and intelligence may be classified at levels that prohibit encoding into model weights for use across

contexts; (3) RAG retrieves current data at inference time, ensuring answers reflect current operational picture; (4) SJA review for fine-tuning on operational corpora is a significant process overhead; risks of RAG: retrieval quality depends on corpus freshness and embedding quality — mitigation: automated corpus refresh pipeline and embedding evaluation. Must cite both tempo AND classification constraints.

SA-3. Full credit: environment: isolated staging copy of the workflow, never production; adversarial scenarios: (1) prompt injection via SITREP text containing override instructions; (2) hallucination probe — inputs designed to elicit fabricated maintenance actions; (3) classification boundary test — inject SECRET-level text into a CUI workflow; (4) role bypass — attempt to trigger Ontology write through the SITREP workflow; (5) data exfiltration attempt — prompt designed to include sensitive context in the output; documentation: test case log with input, expected output, actual output, pass/fail, severity rating; remediation required for all CRITICAL findings; release threshold: zero critical findings, all high findings remediated or formally accepted with mitigation plan.

SA-4. Full credit: ownership structure — named AI system owner (typically the program manager or data officer) who is responsible for the system's compliance and performance; review cycle — quarterly review of output quality metrics, annual comprehensive governance review including compliance mapping; compliance mapping — Army CIO Memo (April 2024), DoD RAIMTF (2024), AR 25-2 for data security; governance review failure procedure — system placed on conditional status with 30-day remediation period; if remediation fails, system is suspended pending command decision; decommission plan must be in place before deployment is approved. Full credit requires all four components.

SA-5. Full credit: root cause analysis — circuit breaker on the write agent was not configured to halt on data quality anomalies; monitoring did not flag the write volume increase (more writes than expected = anomaly indicator); human review gate was not required for all writes (or review cadence was too infrequent); output quality monitoring did not compare written records against source SITREP content; failed controls: (1) rate limit should have flagged unusual write volume; (2) circuit breaker should have stopped chain on first detected data anomaly; (3) daily spot-check of written records; architectural changes: mandatory rate limit with human alert on threshold breach; quality sampling pipeline that compares N% of written records to source data; circuit breaker with data anomaly detection (unexpected field values); all write operations to staging table first with automated data quality check before promotion to production Ontology.

---

*USAREUR-AF Operational Data Team TM-50H Post-Test | Version 1.0 | March 2026*